# Auxiliary problem principle and inexact variable metric forward-backward algorithm for minimizing the sum of a differentiable function and a convex function

Jean-Philippe Chancelier

# Auxiliary problem principle and inexact variable metric forward-backward algorithm for minimizing the sum of a differentiable function and a convex function

Jean-Philippe Chancelier*

August 12, 2015

**Abstract**

In view of the minimization of a function which is the sum of a differentiable function $f$ and a convex function $g$ we introduce descent methods which can be viewed as produced by inexact auxiliary problem principle or inexact variable metric forward-backward algorithm. Assuming that the global objective function satisfies the Kurdyka-Łojasiewicz inequality we prove the convergence of the proposed algorithm extending results of [5] by weakening assumptions found in previous works.

## 1 Introduction

We revisit the algorithms studied in [5] for the minimization of a function which is the sum of a differentiable function $f$ and a convex function $g$. For that purpose we use the Auxiliary Problem Principle (A.P.P.) which was developed in [6]. It allows to find the solution of an optimization problem by solving a sequence of problems called auxiliary problems and as such gives a general framework which can describe a large class of optimization algorithms. One of the basic algorithm which can be obtained through the A.P.P. is the so-called Forward-Backward (F.B.) algorithm [4]. The convergence of the F.B. algorithm has been recently established for nonconvex functions $f$ and $g$ in [3] under the assumption that that the function $f$ is Lipschitz

*CERMICS, École des Ponts ParisTech, 6 & 8 av. B. Pascal, F-77455 Marne-La-Vallée, France

differentiable and that the global objective function satisfies the Kurdyka-Łojasiewicz (KL) inequality [10]. It is of importance to note that given the KL assumption it is possible to prove the convergence of an inexact F.B. algorithm. Inexact F.B. or Inexact A.P.P. means that the auxiliary optimization problems which are iteratively solved can be solved approximately, the approximation will goes to zero as the algorithm converges. Moreover, as pointed in [3], the KL inequality holds for a wide class of functions.

In [5], the authors study a potential way to accelerate the inexact F.B. algorithm through variable metric strategy in the context of nonconvex function. In this article, simplifying the original proof of [3] we can prove the convergence of the inexact variable metric F.B. algorithm of [5] under weaker assumptions.

## 2 Preliminaries

We recall here some standard definitions from variational analysis following [14, 3]. The Euclidean scalar product of $\mathbb{R}^m$ and its corresponding norm are respectively denoted by $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$. For a given positive definite matrix $A$ we denote by $\langle \cdot, \cdot \rangle_A$ and $\|\cdot\|_A$ the scalar product of $\mathbb{R}^m$ and its corresponding norm defined for all $(x, y) \in \mathbb{R}^m \times \mathbb{R}^m$ by

$$\langle x, y \rangle_A = \langle Ax, y \rangle \quad \text{and} \quad \|x\|_A = \langle Ax, x \rangle^{\frac{1}{2}} \ .$$

If $F : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ is a point-to-set mapping its graph is defined by

$$\operatorname{Graph} F \stackrel{\text{def}}{=} \{(x, y) \in \mathbb{R}^m \times \mathbb{R}^m : y \in F(x)\} \ ,$$

while its domain is given by

$$\operatorname{dom} F \stackrel{\text{def}}{=} \{x \in \mathbb{R}^m : F(x) \neq \emptyset\} \ .$$

Similarly, the graph of a real-extended-valued function $\psi : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is defined by

$$\operatorname{Graph} \psi \stackrel{\text{def}}{=} \{(x, s) \in \mathbb{R}^m \times \mathbb{R} : s = \psi(x)\} ,$$

and its domain by

$$\operatorname{dom} \psi \stackrel{\text{def}}{=} \{x \in \mathbb{R}^m : \psi(x) < +\infty\} \ .$$

The epigraph of $\psi$ is defined as usual as

$$\operatorname{epi} \psi \stackrel{\text{def}}{=} \{(x, \lambda) \in \mathbb{R}^m \times \mathbb{R} : \psi(x) \leq \lambda\} .$$

When $\psi$ is a proper function, i.e. when $\operatorname{dom}\psi \neq \emptyset$ , the set of its global minimizers, possibly empty, is denoted by

$$\operatorname{argmin}\psi \stackrel{\text{def}}{=} \{x \in \mathbb{R}^m : \psi(x) = \inf\psi\} \ .$$

The *level set* of $\psi$ at height $\delta \in \mathbb{R}$ is $\operatorname{lev}_{\leq\delta}\psi \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n : \psi(x) \leq \delta\}$. The notion of subdifferential plays a central role in the following theoretical and algorithm developments. For each $x \in \operatorname{dom}\psi$, the *Fréchet subdifferential* of $\psi$ at $x$, written $\hat{\partial}\psi(x)$, is the set of vectors $v \in \mathbb{R}^m$ which satisfy

$$\liminf_{\substack{y\neq x\\ y\to x}} \frac{1}{\|x-y\|}\left(\psi(y) - \psi(x) - \langle v, y-x\rangle\right) \geq 0.$$

When $x \notin \operatorname{dom}\psi$, we set $\hat{\partial}\psi(x) = \emptyset$. The limiting processes used in an algorithmic context necessitate the introduction of the more stable notion of limiting-subdifferential (or simply subdifferential) of $\psi$. The subdifferential of $\psi$ at $x \in \operatorname{dom}\psi$, written $\partial\psi(x)$, is defined as follows

$$\partial\psi(x) \stackrel{\text{def}}{=} \left\{v \in \mathbb{R}^m : \exists x_n \to x, \psi(x_n) \to \psi(x), v_n \in \hat{\partial}\psi(x_n) \to v\right\} \ .$$

It is straightforward to check from the definition the following closedness property of $\partial\psi$: Let $(x_n, v_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}^m \times \mathbb{R}^m$ such that $(x_n, v_n) \in \operatorname{Graph}\partial\psi$ for all $n \in \mathbb{N}$. If $(x_n, v_n)$ converges to $(x, v)$, and $\psi(x_n)$ converges to $\psi(x)$ then $(x, v) \in \operatorname{Graph}\partial\psi$. These generalized notions of differentiation give birth to generalized notions of critical point. A necessary (but not sufficient except when $\psi$ is convex) condition for $x \in \mathbb{R}^m$ to be a minimizer of $\psi$ is

$$0 \in \partial\psi(x)\,. \tag{1}$$

A point that satisfies (1) is called limiting-critical or simply critical.

The derivative of a differentiable function $\psi$ is *strongly monotone* with constant $a$, if it exists $a > 0$ such that

$$\text{for all, } x,\, y \in \mathbb{R}^m \quad \langle \nabla\psi(x) - \nabla\psi(y), x-y\rangle \geq a\,\|x-y\|^2\ . \tag{2}$$

**Remark 1** *If a differentiable and* convex *function $\psi$ satisfy* (2)*, then for all $x$, $y \in \mathbb{R}^m$ we have*

$$D_\psi(y,x) \stackrel{\text{def}}{=} \psi(y) - \psi(x) - \langle\nabla\psi(x), y-x\rangle \geq \frac{a}{2}\,\|x-y\|^2\ . \tag{3}$$

*The function $D_\psi(y,x)$ is called the* Bregmann distance *associated to function $\psi$.*

The derivative of a differentiable function $\psi$ is *Lipschitz* with constant $L$ (or $L$-Lipschitz), if it exists $L > 0$ such that,

$$\|\nabla\psi(x) - \nabla\psi(y)\| \leq L \|x - y\| \ . \tag{4}$$

**Remark 2** *Note, that, thanks to the Lemma 1 (see for example [12, 3.2.12]) when the derivative of a function $\psi$ is L-Lipschitz then we have*

$$\text{for all, } x,\ y \in \mathbb{R}^m \quad D_\psi(y, x) \leq \frac{L}{2} \|x - y\|^2 \ . \tag{5}$$

**Lemma 1 (Descent Lemma)** *Let $\psi : \mathbb{R}^m \to \mathbb{R}$ be a function and $C$ a convex subset of $\mathbb{R}^m$ with nonempty interior. Assume that $\psi$ is $\mathcal{C}^1$ on a neighborhood of each point in $C$ and that $\nabla\psi$ is L-Lipschitz continuous on $C$. Then, for any two points $x,\ u \in C$,*

$$\psi(y) \leq \psi(x) + \langle \nabla\psi(x), y - x \rangle + \frac{L}{2} \|x - y\|^2 \ . \tag{6}$$

# 3 Auxiliary Problem Principle and variations on F.B. Algorithm

We consider here the Auxiliary Problem Principle (A.P.P) for a function $h \stackrel{\text{def}}{=} f + g$ where $g : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is a proper lower semicontinuous convex function and $f : \mathbb{R}^m \to \mathbb{R}$ is differentiable. The core step of the A.P.P. algorithm is to consider the solution of the auxiliary problem

$$y \in \operatorname*{argmin}_{y \in \mathbb{R}^m} T_x(y), \quad \text{with} \quad \mathrm{T}_x(y) \stackrel{\text{def}}{=} \langle \nabla f(x), y - x \rangle + D_K(y, x) + g(y) - g(x). \tag{7}$$

where $D_K$ is the Bregman distance (3) associated to a given core function $K$ which is assumed to be differentiable. Starting with $x_0 \in \operatorname{dom} g$ we iterate the core step to build a sequence $(x_n)_{n \in \mathbb{N}}$ with $x_{n+1} \in \operatorname{argmin}_{y \in \mathbb{R}^m} T_{x_n}(y)$ (Note that this sequence will stay in $\operatorname{dom} g$ and with proper choice of the core $K$, the argmin considered in the iteration is reduced to a unique point). Under technical assumptions the constructed sequence will have a cluster point which is a critical point of the function $h$.

The A.P.P is quite versatile and as developed in [6] many different algorithms can be obtained using a proper choice of the core function $K$. The existence and even uniqueness of a solution to Problem (7) can be ensured by proper choice of of the core function. Note also, that the core function can be replaced by a sequence of functionals which may depend on the iterations and over-relaxation or under-relaxation can be introduced in the sequences. The convergence details are given in [6, Theorem 2.1] under convexity assumptions. We do not recall them here, since our purpose is to focus to the inexact version. We just give an example of a possible core choice which leads to the so-called F.B. algorithm.

Suppose that the core function $K$ is chosen as $K(x) = \|x\|^2/(2\gamma)$ then we obtain

$$T_x(y) \stackrel{\text{def}}{=} \frac{1}{2\gamma}\|y - (x - \gamma\nabla f(x))\|^2 + g(y) - g(x)\,. \tag{8}$$

This choice of $T_x$ operator mixed with under-relaxation with parameter $\lambda \in (0,1]$ gives the so-called F.B. algorithm which consists of the iterations:

$$y_n \in \operatorname{prox}_{\lambda,g}(x_n - \gamma\nabla f(x_n)) \quad \text{and} \quad x_{n+1} = (1-\lambda)x_n + \lambda y_n\,. \tag{9}$$

The proximal operator, $\operatorname{prox}_{\lambda,g}$, being defined by

$$\operatorname{prox}_{\lambda,g}(x) = \operatorname*{argmin}_{y\in\mathbb{R}^m}\left(g(y) + \frac{1}{2\lambda}\|y - x\|^2\right)\,. \tag{10}$$

The minimization problem in (10) has a unique solution when $g$ is a proper convex lower semicontinuous function [7, Lemma 4.1.1]. The Variable Metric Forward-Backward algorithm (V.M.F.B.) is obtained when the core function $K$ is set to a weighted norm (or more precisely to a sequence of weighted norms) $\|\cdot\|_A/(2\gamma)$ where $A$ is a positive definite matrix to be properly chosen. This gives rise to extensions of the prox operator with weighted metric [7, Definition 4.1.2].

Before exposing the inexact F.B. or V.M.F.B. algorithm, we recall a basic property which is satisfied by the iterates of the A.P.P. and which will remains valid in the case of an inexact algorithm. We easily check that $T_x(x) = 0$ for all $x \in \mathbb{R}^m$, and therefore for $y^* \in \operatorname{argmin}_{y\in\mathbb{R}^m} T_x(y)$ we necessarily have $T_x(y^*) \leq 0$. The A.P.P algorithm will thus have iterates in the set $\{y \,|\, T_x(y) \leq 0\}$. This last property is a requested assumption when considering inexact algorithms. We have the following simple characterization which combined with assumptions on the core $K$ will ensure the decrease of the main function $h = f + g$ during iterations.

**Lemma 2** *For all $y \in \mathbb{R}^m$ and all $x \in \operatorname{dom} g$ we have*

$$\{y \,|\, T_x(y) \leq 0\} = \{y \,|\, h(y) + D_{K-f}(y,x) \leq h(x)\}\,. \tag{11}$$

*Proof :* Using the definition of the Bregman distance, we obtain the following equivalent expression of the $T_x$ operator $T_x(y) = h(y) + D_{K-f}(y,x) - h(x)$ and the result follows. $\square$

We end this section showing that choosing $y \in \{y' \,|\, T_x(y') \leq 0\}$ and then $z = (1-\lambda)x + \lambda y$ with $\lambda > 0$ (over or under-relaxation) will ensures decreasing values of $h$.

**Lemma 3** *Let $y \in \mathbb{R}^m$ and $x \in \operatorname{dom} g$ be such that $y \in \{y' \,|\, T_x(y') \leq 0\}$. Assume that the derivative of $f$ is L-Lipschitz and $K$ is such that*

$$D_K(y,x) \geq \frac{c}{2} \|y-x\|^2 \tag{12}$$

*where $c$ is a positive real. Then we have, for any $z = (1-\lambda)x + \lambda y$ and $\lambda > 0$,*

$$h(x) \geq h(z) + \frac{\lambda c - L}{2} \|z-x\|^2 \,. \tag{13}$$

*Proof :* We successively have

$$\begin{aligned}
h(z) =& f(z) + g(z) = f(z) + g((1-\lambda)x + \lambda y) \\
\leq& f(z) + (1-\lambda)g(x) + \lambda g(y) && (g \text{ convex}) \\
\leq& f(x) + \langle \nabla f(x), z-x \rangle + \frac{L}{2} \|z-x\|^2 + g(x) && (\text{with } (4)) \\
&+ \lambda(g(y) - g(x)) \\
\leq& h(x) + \frac{L}{2} \|z-x\|^2 + \lambda \left( \langle \nabla f(x), y-x \rangle + g(y) - g(x) \right) \\
\leq& h(x) + \frac{L}{2} \|z-x\|^2 - \lambda D_K(y,x)\,. && (T_x(y) \leq 0)
\end{aligned}$$

We thus obtain $h(x) \geq h(z) + \frac{(\lambda c - L)}{2} \|z-x\|^2$. $\square$

We turn now to the informal presentation of the inexact F.B. algorithm as described in [5]. The ingredients of the algorithm are as follows. The core functions is chosen as $K(x) \stackrel{\text{def}}{=} (1/2) \|x\|_A^2$ where the given positive definite matrix $A$ is changed during the iterations. Under relaxation is used. The minimization step $y \in \operatorname{argmin}_{y \in \mathbb{R}^m} T_x(y)$ is replaced by a partial minimization. We choose $y \in \{y \,|\, T_x(y) \leq 0\}$ and such that it exists $v \in \partial h(y)$ such that $\|v\| \leq \tau' \|x-y\|$.

## 4 Kurdyka-Łojasiewicz properties

In order to prove the convergence of the inexact F.B. algorithm in the non-convex or non-strongly-monotone case we will use as in [5, 3] the Kurdyka-Łojasiewicz property assumption that we describe here. The main result of this section is Theorem 5 which is the same as [3, Theorem 2.9] with a proof based on a simpler Lemma 4 which enables us to easily take into account relaxation in the proposed algorithm as given in Corollary 6. The Kurdyka-Łojasiewicz property was originally developed in [11, 10, 8]. It was first used in gradient methods in [1] to prove the convergence of descent iterations.

In this section, $a$ and $b$ are fixed positive constants and $h : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is a given proper lower semicontinuous function. For a fixed $x^* \in \mathbb{R}^m$, the notation $h_{x^*}$ denotes the function $h_{x^*}(\cdot) \stackrel{\text{def}}{=} h(\cdot) - h(x^*)$ and $[d < h < e]$ denotes the set $\{x \in \mathbb{R}^m : d < h(x) < e\}$. The following definition is taken from [2] as used in [3].

**Definition 3** *(Kurdyka-Łojasiewicz property) The function $h : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is said to have the Kurdyka-Łojasiewicz property (KL property) at $x^* \in \operatorname{dom} \partial h$ if there exist $\eta \in (0, +\infty]$, a neighborhood $U$ of $x^*$ and a continuous concave function $\phi : [0, \eta) \to \mathbb{R}^+$ such that:*

1. *$\phi(0) = 0$,*
2. *$\phi$ is $C^1$ on $(0, \eta)$,*
3. *for all $s \in (0, \eta)$, $\phi'(s) > 0$,*
4. *for all $x$ in $U \cap [h(x^*) < h < h(x^*) + \eta]$, the Kurdyka-Łojasiewicz inequality holds*

$$\phi'(h(x) - (x^*)) \operatorname{dist}(0, \partial h(x)) \geq 1 \,. \tag{14}$$

*Proper lower semicontinuous functions which satisfy the Kurdyka-Łojasiewicz inequality at each point of $\operatorname{dom} \partial h$ are called KL functions.*

**Assumption 1** *(Localization condition). Let $x^* \in \mathbb{R}^m$ be given, a variable $x \in \mathbb{R}^m$ is said to satisfy assumption $B(U, \eta, \rho)$ if there exists $\rho > 0$ such that*

$$h(x) \in B(h(x^*), \eta), \quad x \in B(x^*, \gamma(x)) \quad \textit{and} \quad B\left(x^*, \gamma(x) + \delta(x)\right) \subset U \tag{15}$$

*where the functions $\gamma$ and $\delta$ are given by*

$$\gamma(x) \stackrel{\text{def}}{=} \rho + \frac{b}{a} \phi\left(h_{x^*}(x)\right) \quad \textit{and} \quad \delta(x) \stackrel{\text{def}}{=} \sqrt{\frac{h_{x^*}(x)}{a}} \,. \tag{16}$$

We start with a technical Lemma. If the function $h$ has the KL property at a point $x^*$, we prove that we can find a neighborhood of $x^*$ which is such that all the values of $y$ satisfying Equations (17) and (18) will stay in the same neighborhood. This Lemma is simpler that the corresponding lemma [3, Lemma 2.6] because we assume that $y$ is such that $h(x^*) < h(y)$. This assumption appears to be sufficient to prove Theorem 5 since equality is treated separately. More precisely:

**Lemma 4** *Assume that the function $h$ has the Kurdyka-Łojasiewicz property at $x^* \in \operatorname{dom} \partial h$ with parameters $(U, \eta, \rho)$ and assume that $x \in \mathbb{R}^m$ satisfy property $B(U, \eta, \rho)$. Let $y \in \mathbb{R}^m$ such that $h(x^*) < h(y)$ satisfying the following inequality*

$$h(y) + a \|x - y\|^2 \leq h(x)\,, \tag{17}$$

*and such that there exists $z \in \partial h(y)$ which satisfy*

$$\|z\| \leq b \|x - y\|\ . \tag{18}$$

*Then, we have that*

$$\|x - y\| \leq \frac{b}{a} \left(\phi(h_{x^*}(y)) - \phi(h_{x^*}(x))\right) \tag{19}$$

*and $y$ satisfy property $B(U, \eta, \rho)$.*

*Proof :* We first show that we can use the KL property at $x^*$ with $y$. Using the fact that $y$ satisfy Equation (17) and that $h(x^*) < h(y)$ we obtain successively that $h(x^*) < h(y) < h(x^*) + \eta$ and

$$\|x - y\| \leq \sqrt{\frac{h(x) - h(x^*)}{a}} \leq \sqrt{\frac{\eta}{a}}\,. \tag{20}$$

This last equation together with the fact that $x$ satisfy $B(U, \eta, \rho)$ gives us that $y \in B(x^*, \gamma(x) + \delta(x)) \subset U$. We can therefore apply the KL property at $x^*$ with $y$. We proceed as follows, let $z$ be in $\partial h(y)$ and satisfying Equation (18), we have that

$$\operatorname{dist}(0, \partial h(y)) \leq \|z\| \leq b \|x - y\|\ .$$

which, combined with the fact that $y$ is such that $y \in U \cap [h(x^*) < h < h(x^*) + \eta]$ and Equation (14) gives

$$\phi'(h_{x^*}(y))^{-1} \leq b \|x - y\|\ . \tag{21}$$

Now, using the concavity of the function $\phi$ we have

$$\phi(h_{x^*}(y)) - \phi(h_{x^*}(x)) \geq \phi'(h_{x^*}(y))(h_{x^*}(y) - h_{x^*}(x))\,. \tag{22}$$

Using the fact that $\phi' > 0$, Equation (22) can be rewritten as

$$(h_{x^*}(y) - h_{x^*}(x)) \leq (\phi(h_{x^*}(y)) - \phi(h_{x^*}(x)))\phi'(h_{x^*}(y))^{-1}$$

(using Equation (21))

$$\leq (\phi(h_{x^*}(y)) - \phi(h_{x^*}(x)))b\,\|x - y\|\ . \tag{23}$$

Using Equation (17), Equation (23) and the inequality $\sqrt{uv} \leq (u+v)/2$ we successively obtain

$$\begin{aligned}\|x - y\| &\leq \left(\frac{h_{x^*}(y) - h_{x^*}(x)}{a}\right)^{1/2} \\ &\leq \left((\phi(h_{x^*}(y)) - \phi(h_{x^*}(x)))\frac{b}{a}\,\|x - y\|\right)^{1/2} \\ &\leq \frac{1}{2}\left(\frac{b}{a}\,(\phi(h_{x^*}(y)) - \phi(h_{x^*}(x))) + \|x - y\|\right)\,. \end{aligned} \tag{24}$$

We finally rewrite Equation (24) as

$$\|x - y\| \leq \frac{b}{a}\,(\phi(h_{x^*}(y)) - \phi(h_{x^*}(x)))\ . \tag{25}$$

It remains to prove that $y$ satisfy $B(U, \eta, \rho)$. Using the fact that $x \in B(x^*, \gamma(x))$ and Equation (25) we obtain

$$\begin{aligned}\|x^* - y\| &\leq \|x^* - x\| + \|x - y\| \\ &\leq \rho + \frac{b}{a}\phi(h_{x^*}(x)) + \frac{b}{a}\,(\phi(h_{x^*}(y)) - \phi(h_{x^*}(x))) \\ &\leq \rho + \frac{b}{a}\phi(h_{x^*}(y)) = \gamma(y)\,, \end{aligned}$$

which gives $y \in B(x^*, \gamma(y))$. Moreover, the function $\phi$ is non-increasing and $h_{x^*}(y) \leq h_{x^*}(x)$ we thus have $\gamma(y) \leq \gamma(x)$ and also $\delta(y) \leq \delta(x)$ which ensures

$$B(x^*, \gamma(y) + \delta(y)) \subset B(x^*, \gamma(x) + \delta(x)) \subset U\,, \tag{26}$$

and we conclude that $y$ satisfy $B(U, \eta, \rho)$. $\square$

**Remark 4** *Suppose that* $z = (1-\lambda)x + \lambda y$ *with* $\lambda \in ]0,1]$ *and* $(x,y) \in \mathbb{R}^{2m}$ *satisfy the assumption of Lemma 4. Then, using the fact that* $\|z-x\| = \lambda \|y-x\|$*, we obtain that* $z$ *satisfy property* $B(U,\eta,\rho)$*.*

**Assumption 2** *We assume that the sequence* $(x_n)_{n\in\mathbb{N}}$ *satisfies the following conditions:*

$(i)$ *(Sufficient decrease condition). For each* $n \in \mathbb{N}$*,*

$$h(x_{n+1}) + a\,\|x_{n+1} - x_n\|^2 \leq h(x_n); \tag{27}$$

$(ii)$ *(Relative error condition). For each* $n \in \mathbb{N}$*, there exists* $w_{n+1} \in \partial h(x_{n+1})$ *such that*

$$\|w_{n+1}\| \leq b\,\|x_{n+1} - x_n\| \ ; \tag{28}$$

$(iii)$ *(Continuity condition). There exists a subsequence* $(x_{\sigma(n)})_{n\in\mathbb{N}}$ *and* $x^*$ *such that*

$$x_{\sigma(n)} \to x^* \quad \text{and} \quad h(x_{\sigma(n)}) \to h(x^*), \quad \text{as} \quad j \to \infty\,. \tag{29}$$

**Theorem 5** *(Convergence to a critical point [3, Theorem 2.9]) Let* $h : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ *be a proper lower semicontinuous function. Consider a sequence* $(x_n)_{n\in\mathbb{N}}$ *that satisfies Assumption 2. If* $h$ *has the Kurdyka-Łojasiewicz property at the cluster point* $x^*$ *specified in Assumption 2-(iii), then the sequence* $(x_n)_{n\in\mathbb{N}}$ *converges to* $x^*$ *as* $k$ *goes to infinity, and* $x^*$ *is a critical point of* $h$*. Moreover the sequence* $(x_n)_{n\in\mathbb{N}}$ *has a finite length, i.e.*

$$\sum_{n=0}^{+\infty} \|x_{n+1} - x_n\| < +\infty. \tag{30}$$

*Proof :* We first show that we can find $n_0 \in \mathbb{N}$ for which $x_{n_0}$ satisfy assumption $B(U,\eta,\rho)$ where $(\phi, U, \eta)$ are the parameters associated with the KL property of $h$ at $x^*$ given in Assumption 2-(iii). Let $x^*$ be the cluster point of $(x_n)_{n\in\mathbb{N}}$ given by Assumption 2-(iii), since $(h(x_n))_{n\in\mathbb{N}}$ is a nonincreasing sequence (as a direct consequence of Assumption 2-(i)), we deduce that $h(x_n) \to h(x^*)$ and $h(x_n) \geq h(x^*)$ for all integers $k$. Then, since $\phi$ is continuous and such that $\phi(0) = 0$ we also have that the sequences

$\gamma(x_n)_{n\in\mathbb{N}} \downarrow \rho$ and $\delta(x_n)_{n\in\mathbb{N}} \downarrow 0$. We choose $\rho' > 0$ such that $B(x^*, \rho') \subset U$, and fix $\rho = \rho'/3$. Let $n_1 \in \mathbb{N}$ be such that

$$\forall n \geq n_1 \quad \gamma(x_n) \leq \frac{\rho'}{2} \quad \text{and} \quad \gamma(x_n) \leq \min(\frac{\rho'}{2}, a\eta^2)\,. \tag{31}$$

Now, since $x^*$ is a cluster point of the sequence $(x_n)_{n\in\mathbb{N}}$, we can find $n_0 \geq n_1$ such that $x_{n_0} \in B(x^*, \gamma(x_{n_0}))$. For $x_{n_0}$ we have

$$h(x_{n_0}) \in B(h(x^*), \eta), \quad \text{and} \quad B(x^*, \gamma(x_{n_0}) + \delta(x_{n_0})) \subset B(x^*, \rho') \subset U\,, \tag{32}$$

and thus $x_{n_0}$ satisfy $B(U, \eta, \rho)$.

Now suppose that the sequence $(x_n)_{n\in\mathbb{N}}$ is such that $h(x_n) > h(x^*)$ for all $n \in \mathbb{N}$. Then, is now possible to apply recursively Lemma 4 for $k \geq n_0$, to obtain that the sequence $(x_n)_{n\geq n_0}$ has a finite length and thus converges to $\overline{x}$. Since $h$ is lower semicontinuous we obtain $h(\overline{x}) \leq h(x^*)$. If is happens that $h(x_{n_1}) = h(x^*)$, then we have $h(x_n) = h(x^*)$ for all $n \geq n_1$ and using Assumption 2-(i)we also have that $x_n = x_{n_1}$ for $n \geq n_1$ and the sequence thus converges to $x^*$. □

**Assumption 3** *We assume that the sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ satisfy the following conditions:*

*(i) (Sufficient decrease condition). For each $n \in \mathbb{N}$,*

$$h(y_n) + a\,\|y_n - x_n\|^2 \leq h(x_n);$$
$$h(x_{n+1}) + a'\,\|x_{n+1} - x_n\|^2 \leq h(x_n);$$

*(ii) (Relative error condition). For each $n \in \mathbb{N}$, there exists $w_n \in \partial h(y_n)$ such that*

$$\|w_n\| \leq b\,\|y_n - x_n\|\ ; \tag{33}$$

*(iii) (Continuity condition). There exists a subsequence $(x_{\sigma(n)})_{n\in\mathbb{N}}$ and $x^*$ such that*

$$x_{\sigma(n)} \to x^* \quad \textit{and} \quad h(x_{\sigma(n)}) \to h(x^*), \quad \textit{as} \quad j \to \infty\,. \tag{34}$$

*(iv) ($\lambda$-Relaxation condition). The two sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ are linked by*

$$x_{n+1} = (1 - \lambda_n)x_n + \lambda_n y_n\,, \tag{35}$$

*where $(\lambda_n)_{n\in\mathbb{N}}$ is a given sequence of reals such that for all $n \in \mathbb{N}$, $\lambda_n \in [\underline{\lambda}, 1]$ and $\underline{\lambda} > 0$.*

**Corollary 6** *(Convergence to a critical point in the under-relaxation case) Let $h : \mathbb{R}^m \to R \cup \{+\infty\}$ be a proper lower semicontinuous function. Consider two sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ that satisfies Assumption 3. If $h$ has the Kurdyka-Łojasiewicz property at the cluster point $x^*$ specified in Assumption Assumption 3-(iii), then the sequence $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ converge to $x^*$ as $k$ goes to infinity, and $x^*$ is a critical point of $h$. Moreover the sequence $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ have a finite length, i.e.*

$$\sum_{n=0}^{+\infty} \|x_{n+1} - x_n\| < +\infty \quad \textit{and} \quad \sum_{n=0}^{+\infty} \|y_{n+1} - y_n\| < +\infty. \tag{36}$$

*Proof :* The proof is very similar to the proof of Theorem 5. Proceeding as in Theorem 5, it is possible to find $n_0 \in \mathbb{N}$ for which $x_{n_0}$ satisfy assumption $B(U, \eta, \rho)$ where $(\phi, U, \eta)$ are the parameters associated with the KL property of $h$ at $x^*$ given in Assumption 2-(iii). Then to proceed as in Theorem 5 we just have to show that the iterates satisfy assumption $B(U, \eta, \rho)$ which is the case using Remark 4. We thus obtain that the $(x_n)_{n\in\mathbb{N}}$ has a finite length and converges to $x^*$ a critical point of $h$. We now prove the result for the sequence $(y_n)_{n\in\mathbb{N}}$. Using Equation (35) we have that $\|y_n - x_n\| \leq (1/\underline{\lambda}) \|x_{n+1} - x_n\|$ which gives the convergence of the sequence $(y_n)_{n\in\mathbb{N}}$ to $x^*$ when $n$ goes to infinity. Then, the inequality

$$\begin{aligned} \|y_{n+1} - y_n\| &\leq \|y_{n+1} - x_{n+1}\| + \|x_{n+1} - x_n\| + \|y_n - x_n\| \\ &\leq \frac{1}{\underline{\lambda}} \|x_{n+2} - x_{n+1}\| + (\frac{1}{\underline{\lambda}} + 1) \|x_{n+1} - x_n\| \end{aligned} \tag{37}$$

gives the finite length property for the sequence $(y_n)_{n\in\mathbb{N}}$. □

# 5 Inexact variable metric forward-backward algorithm

We turn now to the Inexact Variable Metric Forward-Backward (Inexact V.M.F.B.) algorithm studied in [5, 7]. We want to minimize a function $h : \mathbb{R}^m \to ]-\infty, +\infty]$ and assume that $h$ can be split as $h = f + g$ where $f$ is a differentiable function and $g$ is a proper lower semicontinuous and convex function. More precisely, we use the same assumptions as in [5].

**Assumption 4** *The function $h = f + g$, where the functions $f$ and $g$ satisfy the following assumptions:*

*(i) The function $g : \mathbb{R}^N \to ]-\infty, +\infty]$ is proper, lower semicontinuous and convex, and its restriction to its domain is continuous.*

*(ii) The function $f : \mathbb{R}^N \to \mathbb{R}$ is differentiable and its derivative is L-Lipschitz ($L > 0$) on $\operatorname{dom} g$:*

$$\forall (x, y) \in (\operatorname{dom} g)^2 \quad \|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\| \ .$$

*(iii) The function $h = f + g$ is coercive (i.e., $\lim_{\|x\| \to +\infty} h(x) = +\infty$).*

*(iv) The function $h$ satisfies the Kurdyka-Łojasiewicz inequality (Definition 3).*

**Remark 7** *As pointed out in [5], according to Assumption 4, the function $h$ is proper and lower semicontinuous and its restriction to its domain ($\operatorname{dom} h = \operatorname{dom} g$ a nonempty convex set) is continous. Thus, combines with the coecivity of $h$, the level sets of $h$ are compact sets.*

We consider the sequence of A.P.P problems given by

$$T_x^n(y) \stackrel{\text{def}}{=} \langle \nabla f(x), y - x \rangle + D_{K_n}(y, x) + g(y) - g(x). \tag{38}$$

where $(A_n)_{n \in \mathbb{N}}$ is a given sequence of symmetric positive matrices and for all $x \in \mathbb{R}^m$, $K_n(x) = \frac{1}{2\gamma_n} \|x\|^2_{A_n}$ with $(\gamma_n)_{n \in \mathbb{N}}$ a given sequence of reals such that $\gamma_n \in ]0, +\infty)$.

Starting with $x_0 \in \operatorname{dom} g$, the Exact V.M.F.B. algorithm consists in building the sequences $(x_n)_{n \in \mathbb{N}}$ and $(y_n)_{n \in \mathbb{N}}$ as follows:

$$y_n \in \operatorname*{argmin}_y T^n_{x_n}(y) \quad \text{and} \quad x_{n+1} = (1 - \lambda_n) x_n + \lambda_n y_n \,, \tag{39}$$

where $(\lambda_n)_{n \in \mathbb{N}}$ is a sequence of reals such that for all $n \in \mathbb{N}$, $\lambda_n \in [\underline{\lambda}, 1]$ with $\underline{\lambda} > 0$. As such the Exact V.M.F.B. is obtained by applying the A.P.P with weighted metrics and under-relaxation.

Let now $\tau \in ]0, +\infty[$ be given, the Inexact V.M.F.B. Algorithm (as formulated in [5]) iterates starting with $x_0 \in \operatorname{dom} g$ as follows:

$$y_n \in \left\{ y \in \mathbb{R}^m : T^n_{x_n}(y) \leq 0 \right\} \cap \Gamma_{A_n}(x_n) \tag{40}$$

$$x_{n+1} = (1 - \lambda_n) x_n + \lambda_n y_n \,, \tag{41}$$

where $\Gamma_A(x)$ is defined by

$$\Gamma_A(x) \stackrel{\text{def}}{=} \{y \in \mathbb{R}^m : \exists r \in \partial g(y) \quad \text{s.t} \quad \|\nabla f(x) + r\| \leq \tau \|y - x\|_A\}$$

Using previous results we obtain a simple proof of the convergence of the Inexact V.M.F.B algorithm as described by Theorem 8. The proof is simplified when compared to the proof of [5, Theorem 7] and [5, Assumption 3.5] is not needed here. The fact that the Exact V.M.F.B algorithm is a special case of the Inexact V.M.F.B can be found in [5].

**Theorem 8** *(Convergence of the Inexact V.M.F.B algorithm) We consider the sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ obtained using Equations (40) and (41). We assume that*

$$\frac{\underline{\nu}}{2}\|x\|^2 \leq K_n(x) \leq \frac{\overline{\nu}}{2}\|x\|^2 \tag{42}$$

*with $\underline{\lambda\nu} > L$. Then, under assumptions Assumption 4 the sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ satisfy Assumption 3 and the sequence $(x_n)_{n\in\mathbb{N}}$ converges to $x^*$ (Assumption 3-(iii)) and $x^*$ is a critical point of $h$.*

*Proof :* We prove that the sequence $(x_n)_{n\in\mathbb{N}}$ generated by (40) and (41) satisfy Assumption 3. Using Lemma 2 combined with Equation (5) and the choice of $K_n$ we obtain

$$h(y_n) + \frac{1}{2}(\underline{\nu} - L)\|y_n - x_n\|^2 \leq h(x_n)\,. \tag{43}$$

Using Lemma 3 with $c = \underline{\nu}$ and using the fact that $\lambda_n \in [\underline{\lambda}, 1]$, we obtain

$$h(x_{n+1}) + \frac{1}{2}(\underline{\lambda\nu} - L)\|x_{n+1} - x_n\|^2 \leq h(x_n)\,. \tag{44}$$

Assumption 3-(i) is fulfilled with $a = (\underline{\nu} - L)/2$ and $a' = (\underline{\lambda\nu} - L)/2$ which are both positive since $\underline{\lambda\nu} > L$ and $\underline{\lambda} \in ]0, 1]$. Now, using the fact that $y_n \in \Gamma_{A_n}(x_n)$ we obtain $r_n \in \partial g(y_n)$ which is such that

$$\|\nabla f(x_n) + r_n\| \leq \tau \|y_n - x_n\|_{A_n} \tag{45}$$

We thus obtain $w_{n+1} = \nabla f(y_n) + r_n \in \partial h(y_n)$such that

$$\begin{aligned}\|w_{n+1}\| &\leq \|\nabla f(y_n) - \nabla f(x_n)\| + \tau \|y_n - x_n\|_{A_n} \\ &\leq L\|y_n - x_n\| + \tau \|y_n - x_n\|_{A_n} \leq (L + \sqrt{\overline{\nu}}\tau)\|y_n - x_n\|\,.\end{aligned} \tag{46}$$

Assumption 3-(ii)is thus satisfied. It only remains to prove that Assumption 3-(iii)is satisfied, since Assumption 3-(iv)is given by construction.

For Assumption 3-(iii)we first use the fact that the sequence $(x_n)_{n\in\mathbb{N}}$ is bounded as the level sets of $h$ are compact sets (see Remark 7) and $h(x_n)_{n\in N}$ is decreasing. Using the fact that the $h$ is continuous when restricted to its domain we can conclude that Assumption 3-(iii)is satisfied by proceeding as in [3, Theorem 4.2]. $\square$

When $g \equiv 0$, we can reformulate the Inexact V.M.F.B algorithm in order to write $y_n$ as the sum of the solution of the exact algorithm solution and an error $\epsilon_n$. The next Lemma gives the constraint on $\epsilon_n$ which can be deduced from the constraint on $y_n$ given by $T_x^n(y_n) \le 0$.

**Lemma 5** *Suppose that $g \equiv 0$, and for $x \in \mathbb{R}^m$ consider the operator $T_x^n$ given by Equation (38). The the set $T_x^{\le 0} \stackrel{\text{def}}{=} \{y \in \mathbb{R}^m : T_x^n(y) \le 0\}$ has the following equivalent representation*

$$T_x^{\le 0} = \left\{ x - \gamma_n A_n^{-1} \nabla f(x) + \epsilon \quad \textit{with} \quad \|\epsilon\|_{A_n} \le \gamma_n \|\nabla f(x)\|_{A_n^{-1}} \right\} \tag{47}$$

*Proof :* The proof is straightforward and left to the reader. Note that $x - \gamma_n A_n^{-1} \nabla f(x)$ is the unique element in $\operatorname{argmin}_{y'} T_x^n(y')$. $\square$

**Remark 9** *Using Lemma 5 we can reformulate Equations (40) and (41) as follows. $x_n \in \mathbb{R}^m$ being given, choose $\epsilon_n \in \mathbb{R}^m$ such that*

$$\begin{aligned} &\|\epsilon_n\|_{A_n} \le \gamma_n \|\nabla f(x_n)\|_{A_n^{-1}} \\ &\textit{and} \quad \|\nabla f(x_n)\| \le \tau \left\| \epsilon_n - \gamma_n A_n^{-1} \nabla f(x_n) \right\|_{A_n}, \end{aligned}$$

*and update $y_n$ and $x_{n+1}$ with*

$$\begin{aligned} y_n &= x_n - \gamma_n A_n^{-1} \nabla f(x_n) + \epsilon_n \\ x_{n+1} &= (1 - \lambda_n) x_n + \lambda_n y_n \,. \end{aligned}$$

### 5.1 An application to the inexact averaged projection algorithm

In [3, Theorem 3.5] an algorithm is proposed in order to solve a nonconvex feasibility problem and our aim in this section is to derive from Theorem 5 a similar algorithm and its convergence proof. We recall form [3] the context and properties which are used in the algorithm presentation. A closed subset $F$ of $\mathbb{R}^m$ is called *prox-regular* if its projection operator $P_F$ is single-valued around each point $x$ in $F$ (see [13, 3] and references therein). Now, let $F_1, \dots F_p$ be nonempty closed semi-algebraic (See [3] and below), prox-regular subsets of $\mathbb{R}^m$ such that $\cap_{i=1}^p F_i \neq \emptyset$. A classical approach to the problem of finding a common point to the sets $F_1, \dots F_p$ is to find a global minimizer of the function $f : \mathbb{R}^m \to [0, +\infty)$ defined by

$$f(x) \stackrel{\text{def}}{=} \frac{1}{2} \sum_{i=1}^{p} \operatorname{dist}(x, F_i)^2 \tag{48}$$

where $\operatorname{dist}(\cdot, F_i)$ is the distance function to the set $F_i$. When $F$ is prox-regular the function $g(x) = \frac{1}{2} \operatorname{dist}(x, F)^2$ have the following properties [13, 9].

**Theorem 10** *([13]) Let $F$ be a closed prox-regular set. Then, for each $\overline{x}$ in $F$ there exists $r > 0$ such that:*

**(a)** *The projection $P_F$ is single-valued on $B(\overline{x}, r)$,*

**(b)** *the function $g$ is $C^1$ on $B(\overline{x}, r)$ and $\nabla g(x) = x - P_F(x)$,*

**(c)** *the gradient mapping $\nabla g$ is 1-Lipschitz continuous on $B(\overline{x}, r)$.*

The function $f$ given by Equation (48) is semi-algebraic, because the distance function to any nonempty semi-algebraic set is semi-algebraic. This implies in particular that $f$ is a KL function.

We are thus in a situation where the function $f$ has a 1-Lipschitz continuous gradient in a neighborhood of $\overline{x} \in \cap_{i=1}^p F_i$ and is a KL function. Moreover we know that sequences which satisfy Assumption 2 will stay in a neighborhood of $x^*$ specified in Assumption 2-(iii). Then, using Theorem 5, applied to $h = f + g$ with $g \equiv 0$ will ensure the convergence of sequences satisfying Assumption 2 when $x_0$ is sufficiently close to $\cap_{i=1}^p F_i$ and

**Theorem 11** *the sequences $(x_n)_{n \in \mathbb{N}}$, $(y_n)_{n \in \mathbb{N}}$ and $(\epsilon_n)_{n \in \mathbb{N}}$ generated by the following iterates*

$$\begin{aligned} y_n &= x_n - \gamma_n A_n^{-1} R(x_n) + \epsilon_n \\ x_{n+1} &= (1 - \lambda_n) x_n + \lambda_n y_n \,. \quad (\lambda_n \in [\underline{\lambda}, 1], \underline{\lambda} > 0) \end{aligned}$$

*where $\epsilon_n \in \mathbb{R}^m$ is chosen such that $\|\epsilon_n\|_{A_n} \leq \gamma_n \|R(x_n)\|_{A_n^{-1}}$ and $\|\nabla R(x_n)\| \leq \tau \left\|\epsilon_n - \gamma_n A_n^{-1} \nabla R(x_n)\right\|_{A_n}$ with $R : \mathbb{R}^m \rightrightarrows \mathbb{R}$ defined by:*

$$R(x) \stackrel{\text{def}}{=} \sum_{i=1}^{p} (x - P_{F_i}(x)) \, .$$

*are such that the sequence $(x_n)_{n\in\mathbb{N}}$ converges to an element $x^* \in \cap_{i=1}^p F_i$ if $x_0$ is sufficiently close to $\cap_{i=1}^p F_i$ and if the sequences $(K_n)_{n\in\mathbb{N}}$ is such that $\frac{\nu}{2}\|x\|^2 \leq K_n(x) \leq \frac{\overline{\nu}}{2}\|x\|^2$ and $\underline{\lambda\nu} > p$.*

*Proof :* Using Remark 9, the iterates considered in Theorem 11 are similar to the iterates of the inexact V.M.F.B algorithm (Equations (40) and (41)) applied to $h = f + g$ with $g \equiv 0$ and where $\nabla f$ is replaced by $R$. Note that $R$ is not uniquely defined, but it will be when restricted to a neighborhood of $\cap_{i=1}^p F_i$. The conclusion of Theorem 11 will follow from Theorem 8 if we can prove that Assumptions 4 are fulfilled. In fact we only need to have Assumptions 4 in a closed subset of $\mathbb{R}^m$ as noted in [3, Remark 3.3]. We proceed as follows, as it was shown in the proof of Theorem 5 or Lemma 4, if $x_n$ is in a neighborhood of a point $\overline{x}$ then the iterates will stay in the same neighborhood. Using Theorem 10 we can shrink the neighborhood to obtain that $R$ is single-valued and coincide with $\nabla f$ which is $p$-Lipschitz continuous on the selected neighborhood of $\overline{x}$. Thus in neighborhood of $\overline{x}$ the iterates considered in Theorem 11 coincide with the iterates of the inexact V.M.F.B for a function $h = f$ which satisfy Assumptions 4 in a neighborhood of $\overline{x}$ except the coercivity assumption. But since the sequence will stay in a neighborhood of $\overline{x}$ the coercivity assumption is not necessary and we can conclude using Theorem 8. □

**Remark 12** *Even if we chose $\underline{\lambda} = 1$ and $K_n(x) = \|x\|^2 / (2\gamma)$ we do not exactly recover the same algorithm as in [3, Theorem 3.5]*

# 6 Conclusion

In this paper we have recalled the Auxiliary Problem Principle (A.P.P.). It allows to find the solution of an optimization problem by solving a sequence of problems called auxiliary problems and as such gives a general framework which can describe a large class of optimization algorithms. Being able to solve the auxiliary problems not exactly without breakdown in the global algorithm is an important issue. Assuming that the global function to be minimized satisfies the Kurdyka-Łojasiewicz (KL) inequality open the door to such inexact auxiliary problems. Inexact algorithms were developed in [3]. In this paper we have studied the inexact variable metric Forward-Backward algorithm of [5] and proved its convergence under weaker assumptions than in the original work. We have given an application of this algorithm to the inexact averaged projection algorithm [3].

## References


[1] P.-A. Absil, R. Mahony, and B. Andrews. Convergence of the iterates of descent methods for analytic cost functions. *SIAM J. Optim.*, 16(2):531–547, 2005.

[2] Hédy Attouch, Jérôme Bolte, Patrick Redont, and Antoine Soubeyran. Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-łojasiewicz inequality. *Math. Oper. Res.*, 35(2):438–457, 2010.

[3] Hedy Attouch, Jérôme Bolte, and Benar Fux Svaiter. Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods. *Math. Program.*, 137(1-2, Ser. A):91–129, 2013.

[4] George H.-G. Chen and R. T. Rockafellar. Convergence rates in forward-backward splitting. *SIAM J. Optim.*, 7(2):421–444, 1997.

[5] Emilie Chouzenoux, Jean-Christophe Pesquet, and Audrey Repetti. Variable metric forward-backward algorithm for minimizing the sum of a differentiable function and a convex function. *J. Optim. Theory Appl.*, 162(1):107–132, 2014.

[6] G. Cohen. Auxiliary problem principle and decomposition of optimization problems. *J. Optim. Theory Appl.*, 32(3):277–305, 1980.

[7] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Convex analysis and minimization algorithms. II*, volume 306 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1993. Advanced theory and bundle methods.

[8] Krzysztof Kurdyka. On gradients of functions definable in o-minimal structures. *Ann. Inst. Fourier (Grenoble)*, 48(3):769–783, 1998.

[9] A. S. Lewis, D. R. Luke, and J. Malick. Local linear convergence for alternating and averaged nonconvex projections. *Found. Comput. Math.*, 9(4):485–513, 2009.

[10] S. Łojasiewicz. Une propriété topologique des sous-ensembles analytiques réels. In *Les Équations aux Dérivées Partielles (Paris, 1962)*, pages 87–89. Éditions du Centre National de la Recherche Scientifique, Paris, 1963.

[11] Stanislas Łojasiewicz. Sur la géométrie semi- et sous-analytique. *Ann. Inst. Fourier (Grenoble)*, 43(5):1575–1595, 1993.

[12] J. M. Ortega and W. C. Rheinboldt. *Iterative solution of nonlinear equations in several variables*, volume 30 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2000. Reprint of the 1970 original.

[13] R. A. Poliquin, R. T. Rockafellar, and L. Thibault. Local differentiability of distance functions. *Trans. Amer. Math. Soc.*, 352(11):5231–5249, 2000.

[14] R.T. Rockafellar and R. Wets. *Variational Analysis*, volume 317. Springer Verlag, 1998.